A MATHEMATICAL MODEL FOR UNIFORM DISTRIBUTION OF VOTERS PER CONSTITUENCIES

Tomislav Marošević
Department of Mathematics, J. J. Strossmayer University, Osijek
Trg Lj. Gaja 6, HR-31000 Osijek, Croatia
E-mail: tomislav.marosevic@mathos.hr

Kristian Sabo
Department of Mathematics, J. J. Strossmayer University, Osijek
Trg Lj. Gaja 6, HR-31000 Osijek, Croatia
E-mail: ksabo@mathos.hr

Petar Taler
Department of Mathematics, J. J. Strossmayer University, Osijek
Trg Lj. Gaja 6, HR-31000 Osijek, Croatia
E-mail: petar@mathos.hr

Abstract

This paper presents two different approaches on the basis how to generate constituencies. The first one is based on cluster analysis by means of which approach can get compact constituencies having an approximately equal number of voters. An optimal number of constituencies can be obtained by using this method. The second approach is based on partitioning the country into several areas with respect to territorial integrity of bigger administrative units. The units obtained in this way will represent constituencies which do not necessarily have to have an approximately equal number of voters. Each constituency is associated with a number of representatives that is proportional to its number of voters, so the problem is reduced to the integer approximation problem. Finally, these two approaches are combined and applied on the Republic of Croatia.

Key words: *Data Clustering, Clustering, Apportionment Method, Constituencies*

1. INTRODUCTION

According to the current Act on Elections of Representatives to the Croatian Parliament (Narodne novine 116/99, 109/00, 53/03, 69/03-Revised text, 167/03, 44/06, 19/07, 20/09, 145/2010), the Republic of Croatia is divided into ten multi-member electoral districts with a proportional electoral system. Fourteen representatives are elected in each district from the electoral lists. Article 36 in the aforementioned Act reads: “Constituency is determined by the Act on constituencies for the election of representatives to the Croatian Parliament (Narodne novine 116/99) so that the number of voters in constituencies must not vary by more than ±5%. When defining constituency, the Act established the area of districts, cities, and municipalities established by the Act must be taken into account as much as possible.”

Table 1: The number of voters in each constituency in 2007 and 2011.

Constituency	2007		2011	
	Voters Q_j	Deviation from the average number of voters ($\bar{Q} = 382473.1$) $100\frac{Q_j - \bar{Q}}{\bar{Q}}\%$	Voters Q_j	Deviation from the average number of voters ($\bar{Q} = 382436.3$) $100\frac{Q_j - \bar{Q}}{\bar{Q}}\%$
I	361.236	-5.55	358.750	-6.19
II	399.648	4.49	403.716	5.56
III	366.005	-4.31	346.332	-9.44
IV	335.091	-12.39	333.927	-12.68
V	372.163	-2.70	367.654	-3.86
VI	356.575	-6.77	352.471	-7.84
VII	403.812	5.58	413.148	8.03
VIII	385.594	0.82	358.376	0.76
IX	428.590	12.06	440.597	15.21
X	416.017	8.77	422.392	10.45

Table 1 show the number of voters for each constituency in the election years 2007 and 2011. The above data clearly show that there were significantly larger variances in parliamentary elections in 2007 and 2011 than those permitted by the law. The Constitutional Court of the Republic of Croatia warned about this phenomenon in late 2010.

The problem of determining constituencies continues to be a topic of interest in the scientific and technical literature, where one can find algorithms to generate the distribution of constituencies based on various heuristic approaches (Bozkaya et al., 2003; Grilli, 1999; Ricca et al., 2008a,b).

In this paper, we intend to present two different approaches known in the literature, based on which it is possible to generate constituencies. The first approach, given in detail in (Sabo et al., 2012), is based on cluster analysis and presents a modified form of the model (Hess et al., 1965). On the basis of this model, it is possible to generate an optimal electoral constituency that has the property of equally distributing the number of voters among constituencies. In doing so, legally prescribed rules under which the constituencies should have approximately the same number of voters are respected. Furthermore, only geographically close territorial units will enter a constituency.

The second approach is based on the division of the country into several natural units with respect to the territorial integrity of counties. Natural units obtained in this way will represent constituencies, which do not necessarily have to have an approximately equal number of voters. Each constituency is associated with a number of seats that is proportional to the number of voters in that unit, and the problem is reduced to the integer approximation problem. This approach is known as the "Apportionment Methods" (Grilli et al., 1999).

These two approaches can be combined such that first the application of cluster analysis is used to determine an appropriate number of constituencies which have a nearly equal number of voters. Next, the optimal constituencies obtained in this way are then approximated to the nearest natural areas which respect the integrity of the counties, and the number of representatives is determined for each of them on the basis of the Apportionment Methods.

This paper is organized as follows. In Section 2 a mathematical model (Sabo et al., 2012) based on cluster analysis is described together with the algorithm for solving it. Determination of the appropriate number of constituencies in the Republic of Croatia is particularly considered. In Section 3 a brief overview of Apportionment Methods is given. In Section 4 results of the two approaches in the particular example referring to the Republic of Croatia are illustrated.

2. METHOD FOR DETERMINING CONSTITUENCIES WITH APPROXIMATELY EQUAL NUMBERS OF VOTERS

Generally, we consider the area of a state organized in m territorial units (cities or districts), which are determined by their geographical location in the so-called Gauss-Krüger coordinate system by the points $a_i = (x_i, y_i),\ i = 1, \dots, m$. Furthermore, we assume that territorial unit a_i has q_i voters, and that the total number of voters Q is specified by $\sum_{i=1}^{m} q_i = Q$. We want to share the territory of the country into $k, (1 < k < m)$ constituencies $\pi_1, \dots, \pi_k$, each having the corresponding number of voters $Q_1, \dots, Q_k$, so that: (i) each constituency consists of the territorial units (cities and districts) that are close to each other, which is determined using some distance measure; (ii) the requirement of the first paragraph of Article 36 in the Act on the Election of Representatives to the Croatian Parliament (Narodne novine, 116/99) is satisfied.

The proximity requirement of territorial units of a constituency may be provided by applying cluster analysis using appropriate distance-like function $d: \mathbb{R}^2 \times \mathbb{R}^2 \to [0, +\infty\rangle$ (see e.g. Kogan, 2007). The so-called *least square* (LS) distance-like function[1] $d(x,y)=\|x-y\|^2$ is commonly used in various applications because it is simple and intuitive. Hence, we will use this metric function hereinafter.

We will consider the mathematical model described in (Sabo et al. 2012) where both criteria (i) and (ii) are satisfied. In order to improve the model, it is possible to take into account some other criteria that ensure additional conditions of uniformity, such as: *Socio-economic homogeneity* (Bourjolly, 1981; Bozkaya et al., 2003), *Similarity with the existing structure of constituencies* (Bozkaya et al., 2003), *Area similarity* (Bozkaya et al., 2003), *Keeping major territorial units (e.g. counties)*.

Assume that the number of voters in the two constituencies must not differ by more than p%. It is not hard to show that in this case the number of voters Q_j in any constituency must not differ from the average number of voters per constituency $\frac{Q}{k}$ for more than $p'=\frac{100p}{200+p}$ % (see Sabo et al., 2012).

We will associate every constituency $\pi_1,\dots,\pi_k$ with its corresponding center $c_1,\dots,c_k$. If we use the LS-distance-like function according to (Chen and Peng, 2008) and (Teboulle, 2007) the problem of determining optimal constituencies is reduced to the following optimization problem:

$$\min_{w_{ij}}\sum_{i=1}^{m}\sum_{j=1}^{k}w_{ij}q_i\left\|a_i-\frac{\sum_{s=1}^{m}w_{sj}q_sa_s}{\sum_{s=1}^{m}w_{sj}q_s}\right\|^2. \qquad (1)$$

with the following conditions:

$$\sum_{j=1}^{k}w_{ij}=1,\ \ i=1,\dots,m \qquad (2)$$

$$\sum_{i=1}^{m}\sum_{j=1}^{k}w_{ij}q_i=Q \qquad (3)$$

$$\left(1-\frac{p'}{100}\right)\frac{Q}{k}\le\sum_{i=1}^{m}w_{ij}q_i\le\left(1+\frac{p'}{100}\right)\frac{Q}{k} \qquad (4)$$

$$w_{ij}\in\{0,1\},\ \ i=1,\dots,m,\ \ j=1,\dots,k. \qquad (5)$$

Condition (2) ensures that each territorial unit a_i belongs to precisely one constituency π_j, condition (3) ensures that the voters of every territorial unit are included in some general constituency, condition

[1] LS-distance like-function is defined as the square of ordinary Euclidean distance. In this case, the center of constituency is represented by the point ("centroid, center of gravity") which has the property that the sum of the squared distances to their territorial units is minimal.

(4) ensures the uniformity of the number of voters per constituency for the most p'%, and condition (5) ensures that each territorial unit a_i belongs entirely to only one constituency.

If we assume there is a town or a municipality with population q_{i_0}, which is more than p'% higher than the average size of constituencies, i.e., for whose population q_{i_0} we have

$$q_{i_0} \geq \left(1 + \frac{p'}{100}\right)\frac{Q}{k}, \tag{6}$$

then optimization problem (1) with conditions (2) - (5) does not have a solution. In that case, we must allow that such territorial unit splits in more than one constituency. This is achieved by imposing index set I_0 of those territorial units to which (6) applies, and instead of conditions (5), the following condition is introduced

$$w_{ij} = \begin{cases} [0,1], & i \in I_0 \\ \{0,1\}, & i \in \{1, \dots, m\} \setminus I_0 \end{cases}, \quad j = 1, \dots, k. \tag{7}$$

Optimization problem (1) with conditions (2) - (5) and (7) is a nonlinear global optimization problem with restrictions (Floudas and Gounaris, 2009), with a very large number of variables, and a lot of potential local solutions. Direct solving of this problem is extremely demanding from the numerical aspect. Instead of solving this global optimization problem, motivated by the work of (Ng, 2000), in (Sabo et al., 2012) we have constructed an iterative algorithm that provides a locally optimal solution by using a good initial approximation[2].

Assuming that we launched the algorithm from (Sabo et al., 2012) with a good initial center approximation $c_1, \dots, c_k$, we can claim that the result of algorithm execution provided locally optimal centers of constituencies $c_1^*, \dots, c_k^*$. The constituency has been defined with $\pi_j^* = \{a_i \in \mathcal{A}: w_{ij}^* \neq 0$, $j=1, \dots, k$. It should be noted that two possibilities can occur for $a_i \in \pi_j^*$: if $w_{ij}^*=1$, then the territorial unit a_i is contained entirely in the constituency π_j^*; if w_{ij}^* is a number between 0 and 1, then the constituency π_j^* contains the w_{ij}^*th part of the territorial unit a_i (this option will occur only if $i \in I_0$, where I_0 is the index set of those territorial units for which (6) applies).

According to (Leisch, 2006), our iterative procedure is initiated with e.g. 1000 different randomly generated initial centers, and the one that gives the smallest value of cost function (1) is taken as a

[2] It takes 5 minutes to implement Algorithm 1, which involves determining the initial approximation based on data for the Republic of Croatia with 10 constituencies, on the computer with an Intel Core i5 760 processor and 8GB of RAM. Routines from the GNU Linear Programming Kit (GLPK) library (http://www.gnu.org/s/glpk) are used for linear and integer programming.

solution. Unfortunately there isn't any guarantee that this solution is global, but we can believe that it is optimal with high probability.

3. APPORTIONMENT METHODS

Assume that an electoral system consists of $k \in \mathbb{N}$ constituencies. The total number of seats S should be apportioned among these k electoral districts, so that a number of seats s_j which belongs to the constituency j $(j = 1, \dots, k)$ is determined according to the number of voters Q_j in that constituency, in such a way that s_j is "most proportional" to Q_j, for all $j = 1, \dots, k$. Thus, for given S seats, k constituencies and the corresponding vector of voters $\mathbf{Q} = (Q_1, \dots, Q_k)$, where the total number of voters is $Q = \sum_{j=1}^{k} Q_j$, the apportionment problem is to determine the k – tuple $s = (s_1, \dots, s_k)$ such that there exists a degree of proportionality between the number of seats s_j and the number of voters q_j, for all constituencies $j = 1, \dots, k$, where $\sum_{j=1}^{k} s_j = S,\ s_j \in \mathbb{N}$.

Different proportional electoral methods may be used for apportionment of seats to constituencies according to their number of voters and they depend on different measures of proportionality (Grilli et al., 1999, Marošević et al., 2007).

Proportional electoral formulas can be classified into two groups: quotient methods and divisor methods. Quotient methods are based on quotients, i.e. quotas z_j allocated to each constituency:

$$\lfloor z_j \rfloor \le s_j \le \lceil z_j \rceil, \qquad j = 1, \dots, k.$$

In quotient methods apportionment of seats to constituencies according to the number of their voters is done considering integer parts of quotas, and then the remaining seats are apportioned to those districts which have largest remainders.

Well-known quotient methods are given in Table 2.

Table 2: Quotient methods and their quotas

Quotient Method	Quota
Largest Remainders method (Hamilton method)	(natural quota) $z_j = \frac{Q_j}{Q} \cdot S$
method of Droop quota	$z_j = \frac{Q_j}{Q} \cdot (S + 1)$
method of Imperiali quota	$z_j = \frac{Q_j}{Q} \cdot (S + 2)$

Divisor methods are determined by their increasing sequence of divisors $d(0) < d(1) < d(2) < \ldots < d(S-1)$. For each constituency j and every divisor one computes the ratio between the number of voters q_j and corresponding divisors $d(m),\ m = 0, 1, \ldots, S-1$:

$$\frac{Q_j}{d(0)} > \frac{Q_j}{d(1)} > \cdots > \frac{Q_j}{d(S-1)}, \qquad \forall j = 1, \ldots, k.$$

Then S seats are apportioned to those constituencies which have the S largest ratios.

Well-known divisor methods are given in Table 3.

Table 3: Divisor methods and their divisors

Divisor method	Divisors
Smallest Divisors (i.e. Adams method)	$d(m) = m\ \ (0, 1, 2, \ldots)$
Danish	$d(m) = 1 + 3m\ \ (1, 4, 7, \ldots)$
Harmonic Mean	$d(m) = \dfrac{2m(m+1)}{2m+1}\ \ (0, \approx 1.33, 2.4, \ldots)$
Equal Proportions (i.e. Huntington method)	$d(m) = \sqrt{m(m+1)}\ \ (0, \approx 1.41, \ldots)$
Sainte-Laguë (i.e. Webster method)	$d(m) = 1 + 2m\ \ 1, 3, 5, \ldots)$
modified Sainte-Laguë	(1.4, 3, 5, 7, ...)
d'Hondt (i.e. Jefferson method)	$d(m) = m + 1\ \ (1, 2, 3, \ldots)$
Belgian	$d(m) = 1 + \dfrac{m}{2}\ \ (1, 1.5, 2, 2.5, \ldots)$

In connection with the case of "dividing by zero", let us note that if some divisor method has $d(0) = 0$ (as for methods: Smallest Divisors, Harmonic Mean and Equal Proportions), then by that method each constituency gets one seat in advance, regardless of their number of voters Q_j $(j = 1, \ldots, k)$.

4. CONSTITUENCIES IN THE REPUBLIC OF CROATIA

4.1. Determination of constituencies with approximately equal numbers of voters

The area of the Republic of Croatia is organized into $m = 556$ territorial units (cities or districts). We would like to divide the Croatian territory into $k, (1 < k < m)$ constituencies. The number of voters per constituencies must not differ from each other by more than 5%, i.e., condition (6) with $p' = \frac{100}{41}$ must be satisfied see (Sabo et al., 2012). At the same time, we assume that the number of constituencies k ranges from 2 to 10. For example if we suppose that the number of constituencies is 10, on the basis of model described in Section 2 the optimal division of Croatian territory shown in

Figure 1. Here, each territorial unit belongs to exactly one constituency, except the city of Zagreb, which has to be divided between constituencies I (56.3%), II (42.7%), and VII (1%).

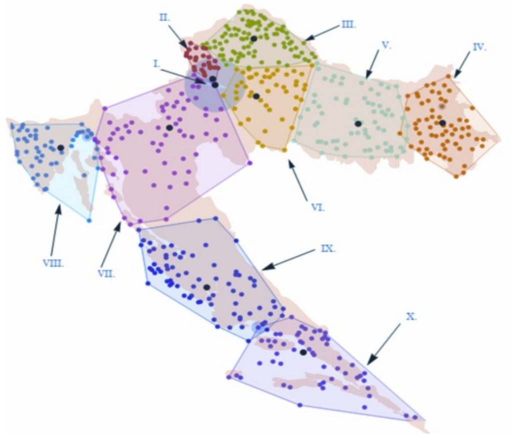

Figure 1: Distribution of the Croatian territory into $k = 10$ *constituencies.*

The question naturally arises as to how many constituencies would be most appropriate for the current geopolitical status of Republic of Croatia. In other words, how many constituencies should there be, and how should they be configurated, such that the compactness of territory is secured, where we need to keep the uniformity of the number of voters per constituencies as it is predicted by the mathematical model.

Automatically determining the number of clusters has been one of the most difficult problems in data clustering processes. In some cases, the number of clusters in a partition is determined by the nature of the problem itself. If the number of clusters in a partition is not given in advance, then it is natural to search for an optimal partition which consists of clusters that are as compact and relatively strongly separated as possible.

The literature attempts to answer this question by using various indices such as the Davies-Bouldin index, Calinski-Harabasz index, or the decrease rate of the cost function (Kogan, 2007; Sabo et al., 2010; Gan et al., 2007). It also makes sense to minimize the number of constituencies that need to share the largest territorial unit - the city of Zagreb. It is usually necessary to combine all indices, and choose a number that indicates the highest index number for the number of constituencies. A higher value of the Calinski-Harabasz index corresponds to a more appropriate number of constituencies. Figure 2a shows that, according to this criterion, we should take 5, 6, 9, or 10 for the number of constituencies. If we look at the value of the cost function (Figure 2b), we see that its rate of decrease is significantly higher up to to 5 and 6 constituencies and then it "flattens" out, meaning that according to this criterion, an appropriate number of constituencies was 5 or 6. We are also going to reduce the division of city of Zagreb by using 5 or 6 constituencies (Figure 2c).

a) Calinski-Harabasz index depending on the number of constituencies

b) Cost function values depending on the number of constituencies

c) Number of constituencies the city of Zagreb is divided into depending on the total number of constituencies

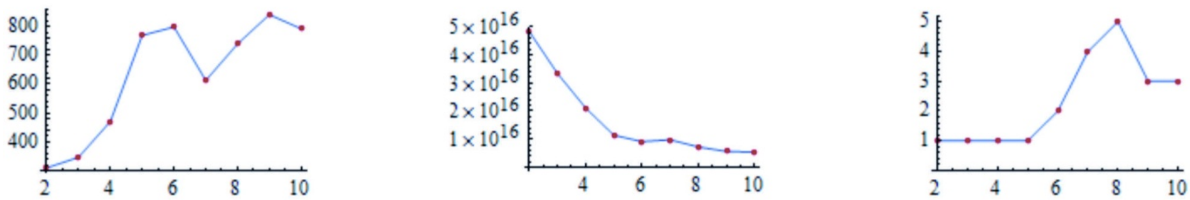

Figure 2: The values of different indices, depending on the number of constituencies.

The application of these indices to the problem of determining the optimal number of constituencies in the Republic of Croatia indicates that the appropriate number of constituencies should be 5, in case we do not want to divide the city of Zagreb, or 6, if we allow its division. Figure 3 shows the division of Croatia into 5 and 6 constituencies. The black dots denote geographic centers of each cluster (constituency).

$k = 5$ constituencies

$k = 6$ constituencies

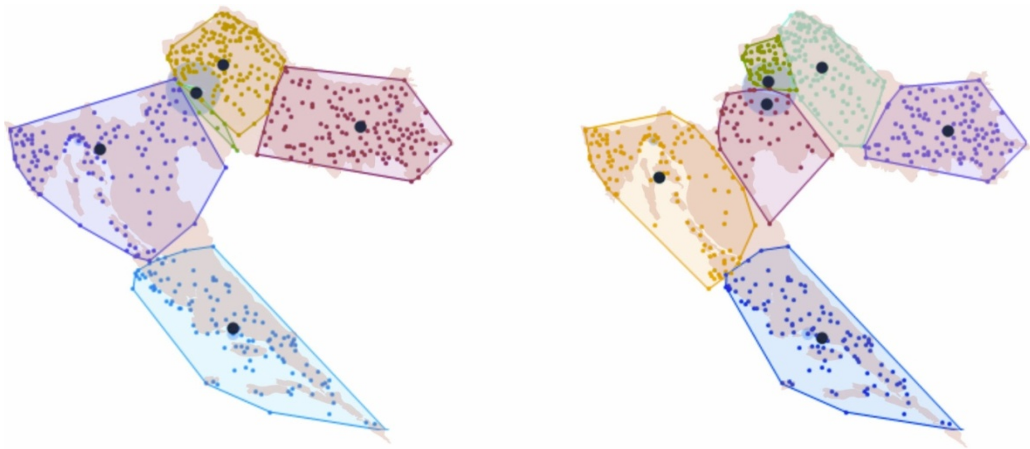

Figure 3: Division into $k = 5$ and $k = 6$ constituencies according to the Integer Approach.

4.2. Application of the Apportionment Methods

In the previous subsection, we concluded that the optimal number of constituencies in the Republic of Croatia is 5 or 6. Suppose that the Croatian territory is divided into five constituencies, but such that the integrity of the counties is preserved. It makes sense to generate a constituency in the way that the areas resulting from Subsection 4.1 are approximated with fairly similar areas, while preserving the integrity of the counties. In this way, we obtain a constituency as shown in Table 3 and Figure 5.

Table 4: Constituencies that preserve the integrity of counties, obtained by approximating constituencies from Subsection 4.1.

Constituency	County	The number of voters on the basis of data used for the 2012 Census
Constituency I	City of Zagreb, Zagreb County	1,013,125
Constituency II	Krapina-Zagorje, Sisak-Moslavina, Varaždin, Koprivnica-Križevci, Bjelovar-Bilogora, Međimurje	761,239
Constituency III	Virovitica-Podravlje, Požega-Slavonia, Brod-Posavina, Osijek-Baranja, Vukovar-Srijem	788,584
Constituency IV	Karlovac, Primorje-Gorski Kotar, Lika-Senj, Istria	677,313
Constituency V	Zadar, Šibenik-Knin, Split-Dalmatia, Dubrovnik-Neretva	851,876

Let us consider the problem of proportional apportionment of seats to constituencies for the data from the previous Table 4, on the basis of the given number of voters Q_j, in the constituencies $j = 1, \dots, 5$. In this way, we have the total number of seats $S = 140$, $k = 5$ constituencies, with the corresponding number of voters Q_j given in the third column. We have applied 11 proportional electoral methods mentioned in Section 3 for apportionment of seats to these five constituencies, and obtained the same results, independently of which proportional electoral method was used. It is the following apportionment of seats to electoral districts:

$$s_1 = 35, \qquad s_2 = 26, \qquad s_3 = 27, \qquad s_4 = 23, \qquad s_5 = 29$$

So, we can conclude that this apportionment indisputably represent proportional apportionment of seats to five electoral districts according to the number of their voters for the data from the previous Table 4.

Figure 4: Constituencies from Table 4 with the corresponding number of representatives, obtained by applying all apportionment methods described in Section 3.

REFERENCES

Act on Elections of Representatives to the Croatian Parliament, Narodne novine 116/99, 109/00, 53/03, 69/03-Revised text, 167/03, 44/06, 19/07, 20/09, 145/2010.

Balinski, M. L. and Young, H.P. (1975), "The quota method of apportionment", *Amer. Math. Monthly*, Vol. 82, pp. 701–729.

Bourjolly, J.M., Laporte, G. and Rousseau, J. M. (1981), "Découpage électoral automatisé á I'le de Montréal", *INFOR*, Vol. 19, pp. 113–124.

Bozkaya, B., Erkut, E. and Laporte, G. (2003), "A tabu search heuristic and adaptive memory procedure for political districting", *European Journal of Operational Research*, pp. 12–26.

Floudas, C.A. and Gounaris, C.E. (2009), "A review of recent advances in global optimization", *Journal of Global Optimization*, Vol. 45, pp. 3–38.

Gan, G., Ma, C. and Wu, J. (2007), *Data Clustering: Theory, Algorithms, and Applications*, SIAM, Philadelphia.

Grilli de Cortona, P., Manzi, C., Pennisi, A., Ricca, F. and Simeone, B. (1999), *Evaluation and optimization of electoral systems*, SIAM Monographs on Discrete Mathematics, Philadelphia.

Hess, S.W., Weaver, J.B., Whelan, H.J. and Zitlau P.A. (1965), "Nonpartisian political redistricting by computer", *Operations Research*, Vol. 13, pp. 998–1006.

Izvješće o nejednakoj težini biračkog glasa u izbornim jedinicama određenima člancima 2. do 11. Zakona o izbornim jedinicama za izbor zastupnika u Zastupnički dom Hrvatskog sabora, Narodne novine 116/99, Narodne novine 142/10.

Kogan, J. (2007), *Introduction to clustering large and high-dimensional data*, Cambridge University Press, Cambridge.

Leisch, F. (2006), "A toolbox for K-centroids cluster analysis", *Computational Statistics & Data Analysis*, Vol. 51, pp. 526–544.

Marošević, T. and Scitovski, R. (2007), "An application of a few inequalities among sequences in electoral systems", *Applied Mathematics and Computation*, Vol. 194, pp. 480–485.

Ng, M.K. (2000), "A note on constrained k-means algorithms", *Pattern Recognition*, Vol. 33, pp. 525–519.

Ricca, F., Scozzari, A. and Simeoni, B. (2008), "Weighted Voronoi region algorithms for political districting", *Mathematical and Computer Modeling*, Vol. 48, pp. 1468–1477.

Ricca, F. and Simeoni, B. (2008), "Local search algorithms for political districting", *European Journal of Operational Research*, pp. 1409–1426.

Sabo, K., Scitovski, R. and Taler, P. (2012), "Ravnomjerna raspodjela broja birača po izbornim jedinicama na bazi matematičkog modela", *Hrvatska i komparativna javna uprava*, Vol 14, pp. 229 – 249.

Sabo, K., Scitovski, R and Vazler, I. (2010), "Grupiranje podataka: klasteri", *Osječki matematički list*, Vol. 10, pp. 149–176.

Sierksma, G. (2002), *Linear and Integer Programming: Theory and Practice*, 2nd Edition, Marcel Dekker, New York.

Teboulle, M. (2007), "A unified continuous optimization framework for center-based clustering methods", *Journal of Machine Learning Research*, Vol. 8, pp. 65–102.